LEVEL II

(2)

ADA079350

DDC FILE COPY

Research Memorandum 70-7

ANALYSIS OF OFFICER PERFORMANCE ON AN EXPERIMENTAL TASK: AUTOMOTIVE INSPECTION

BESRL

U. S. Army
Behavior and Systems Research Laboratory

December 1970

14 BESRL-RM-70-7

16 Army Project Number 1Q062106A722

Officer Prediction d-82

9 Research Memo[redacted] 70-7

6 ANALYSIS OF OFFICER PERFORMANCE ON AN EXPERIMENTAL TASK: AUTOMOTIVE INSPECTION.

10 Edward M. Sait

Louis P. Willemin, Task Leader

12 22

A

Submitted by:
William H. Helme, Chief
Behavioral Evaluation
Research Division

Approved by:
J. E. Uhlaner, Director
U. S. Army Behavior and
Systems Research Laboratory

11 Dec[redacted]70

Research Memorandums are informal reports on technical research problems. Limited distribution is made, primarily to personnel engaged in research for the Behavior and Systems Research Laboratory.

402 177

ANALYSIS OF OFFICER PERFORMANCE ON AN EXPERIMENTAL TASK:
AUTOMOTIVE INSPECTION

The Automotive Inspection Task is one of fifteen situational performance tests developed and administered as part of a large-scale longitudinal research effort in the area of officer leadership. The research was initiated by BESRL in response to recommendations by the Army Scientific Advisory Panel (ASAP) and by the Deputy Chief of Staff for Personnel (DCSPER). The former indicated a need for additional research on the performance and selection of combat officers and suggested that dimensions of such performance might be defined by means of performance exercises within a combat simulation. DCSPER, in view of the increasing complexity of military technology, was interested in determining the feasibility of differential prediction of performance for broad areas of possible officer specialization. The research design of the program incorporates both sets of requirements.

The research is concerned with three broad areas--combat, administrative, and technical. Experimental predictor tests relevant to these areas were administered to 6,900 officers on entrance to active duty in 1958-59, and a revised battery to 4,000 officers on entrance to active duty in 1961-1964. One to two and one-half years after testing, a subsample of 900 of the latter group, six at a time, participated in an exercise at the Officer Evaluation Center (OEC) established for the purpose at Fort McClellan, Alabama. There, in a simulated Military Assistance Advisory Group (MAAG) setting, over a period of three days, a scenario unfolded which eventuated in invasion and guerrilla warfare. The six officers received a series of assignments, first administrative and technical, and then combat. Performance was recorded and rated out of sight of the examinee by cadre who played the parts of MAAG, host nation, and aggressor personnel. Work products were retained for later scoring. The performance records and work products, after analysis to define dimensions of officer performance at the OEC, will serve as criteria for the predictor tests.

The Automotive Inspection Task is one of five in the technical area and was administered on the first day. The examinee was required to perform an inspection of three vehicles (two M-38 jeeps and one M-37 three-quarter ton) all of which were to be brought into good running condition. The examinee was to enter identifying information, deficiencies and shortcomings, and required corrective actions on Equipment Inspection and Maintenance Worksheets. If time allowed, he was also to undertake any of the required repairs that were feasible with the simple tools available. Remaining repairs were to be made by host-nation mechanics, following the examinee's written orders. Relevant manuals were available, and an inexperienced enlisted man was assigned to act as assistant to the examinee. Nearly three hours' time was allowed.

The worksheets prepared by the examinee were the principal basis for scoring, which was carried out using Automotive Inspection Scoring Forms, one for each vehicle. Each form covered seven items of identifying information for which a credit could be given and fourteen prearranged vehicle defects. For a given defect, a symptom credit was given if the major symptom was recorded by the examinee (e.g., "engine won't turn over") but not the cause; a location credit was given for the cause (e.g., "battery ground wire disconnected"); and a correction action credit was given for any of one or more repairs, established in advance as adequate, ordered or made by the examinee. In scoring, actual repair was distinguished from ordering a repair.

Two other scoring forms, the Problem Approach Checklist and the Descriptive Report II, provided subjective evaluations of the examinee by the enlisted assistant. The characteristics evaluated are described below.

OBJECTIVES

The main objective of this analysis was to obtain scores representing the principal behavior dimensions of performance in the test and a score representing overall performance. These scores are to be correlated with scores from the other fourteen tasks to indicate the total structure of leadership behaviors involved in performance of the entire OEC exercise. From the scores on this and the other tasks, criterion scores will be derived to validate the experimental predictor test. Another objective was to evaluate reliability and other characteristics of the scores on the Automotive Inspection Task.

METHOD

SAMPLE

The sample consisted of 733 examinees from the point at which testing procedures were well stabilized (Group 39) through the last group tested (Group 159).

VARIABLES

The variables defined below, except for Importance Ratings, were obtained directly from the scoring forms or derived from data on these forms after initial decisions concerning the weighting and combination of elementary credits. These variables are grouped below by type of scoring form that served as the source document. Certain complex indices are described under "Analysis."

Automotive Inspection Scoring Forms (Objective Scoring Record). The scores defined below, except for Defect Scores, were obtained separately for each vehicle and then summed to provide across-vehicle total scores.

1. Identifying Information

For each vehicle the same seven items representing initial entries required on the examinee's worksheets were scored. The seven items were summed with unit weights to provide an Identifying Information score.

2. Diagnostic Sum

For each defect, one credit was given if only the symptom was detected, two if the location was identified. Thus, a given defect could be scored 0, 1, or 2. Totals across the fourteen defects provided the Diagnostic Sum score for each vehicle.

3. Repair Scores

Ordered Repairs. For each vehicle, the number of defects for which an appropriate repair was requested constituted the vehicle Ordered Repairs score.

Made Repairs. For each vehicle, the number of defects corrected by the examinee or by the assistant under his instruction constituted the vehicle Made Repairs score.

Repair Sum. The Repair Sum for a vehicle was the total of the Ordered Repairs and Made Repairs scores.

4. Defect Scores

For each of the 42 prearranged defects (14 per vehicle), a score was given by adding to the defect diagnostic score one point for corrective action recommended or performed. Making a repair usually required location of the trouble, which was credited with 2 points, resulting in a total score of 3. However, for some items, appropriate repair instructions could be given when only the symptom had been noted. Therefore, adequate instruction to repair the defect could be associated with an item score of either 2 or 3.

5. Total Score

Three trial total scores were obtained for each vehicle and across vehicles (Variables 6, 7, and 8 of Tables 1 and 2) for use in computing special indices (see "Analysis") and in guiding formulation of a final total score. The first total consists of the sum of the defect scores just described, equivalent to the Diagnostic Sum plus the Repair Sum. The second is identical except that a credit of two points instead of one was given for each repair made rather than ordered.

The third score consists of the second total plus the Identifying Information score. On the basis of various considerations, including the statistics for these scores, a final total score was established as described under "Results."

6. Importance Ratings

Three officers and five enlisted men who administered the test each rated on a 10-point scale the extent to which each possible score on the Automotive Inspection Scoring Forms represented an important contribution to the accomplishment of the assigned mission. For each score, the eight ratings were averaged to the nearest whole number. These ratings were considered in item evaluation and in establishing the relative weights of the components of defect and total scores.

<u>Problem Approach Checklist</u> (Judgmental Ratings)

1. Trouble-shooting Approach

A check mark indicated whether, in the judgment of the enlisted assistant, the examinee had a definite plan for trouble-shooting and held to that plan, had such a plan but failed to complete one phase before going on to another, or had no definite plan. The three alternatives were scored 2, 1, and 0.

2. Utilization of Personnel

A check mark by the enlisted assistant indicated whether the examinee made effective use of the assistant (giving him instruction when necessary), used him only for simple tasks (not requiring appreciable instruction), or failed to use him advantageously. Scores of 2, 1, and 0 were given these alternatives.

3. Use of Available Manuals

A check mark by the assistant indicated whether the examinee made efficient use of available manuals, knew the equipment and so did not need to use the manuals, lacked this knowledge but made no use, or spent an excessive time (his or the assistant's) on the manuals. Scoring of these alternatives was 4, 3, 0, and 1.

4. Importance Ratings

Importance ratings were obtained for the alternatives of each of the three scores of the Problem Approach Checklist as for the inspection forms, but on a 20-point scale. These importance ratings were considered in establishing the scoring described above.

Descriptive Report II (Judgmental Ratings)

1. Motivation and Attitude

These were separately rated by the assistant on a 5-step scale (outstanding, excellent, satisfactory, questionable, poor). The steps were assigned the numerical values 5, 4, 3, 2, 1.

2. Factors Considered

The assistant checked those of the ten factors listed below to which he would give most weight if he were evaluating the examinee's overall performance. Then, if the examinee was considered strong on a checked factor, a preprinted "+" was circled; if weak, a "-". Each factor was scored by coding a minus as 0, a plus as 2, and neither as 1.

Bearing and assurance

Effective expression (written or oral)

Keeping cool

Endurance and stamina

Familiarity with the equipment

Following instructions

Extent to which the mission was accomplished

Effective command and control

General impression

Other (to be specified)

ANALYSIS

Item Statistics. P-values were obtained for all unit-level scores--identifying-information entries, symptom and location determinations, and corrective-action alternatives--on the Automotive Inspection Scoring Forms for use in evaluating these scores and for description of examinee performance. Means, standard deviations, and intercorrelations were obtained for the 42 defect (item) scores. These statistics were used for item evaluation and for measurement of internal consistency reliability.

Special Indices. The following four special indices were obtained from other scores in an attempt to measure additional performance characteristics which might be of practical significance.

1. Concentration vs Distribution of Effort

This is a measure of the extent to which total scores, one for each vehicle, tend toward similarity or divergence. Two extreme cases can occur, the first when all scoring points are obtained on one vehicle, the second when vehicle scores are identical. Divergence represents, with some degree of error, concentration of effort on one or two of the vehicles to be inspected at the expense of the other two or one. The Index was formulated as follows:

$$100\sqrt{\frac{3}{2}\left[\frac{A^2 + B^2 + C^2}{(A + B + C)^2} - \frac{1}{3}\right]},$$

where A, B, and C represent respective total scores for each of the three vehicles. The third and more comprehensive of the trial total scores, as defined under variables, was used. Resulting scores can range from 0 for equality of vehicle totals to 100 when all scoring credits were obtained on one vehicle only. The purpose of this score was to measure what might be a general tendency, across several of the tests and in practical situations, to emphasize thoroughness of work as opposed to completion of general overall requirements at the expense of thoroughness.

2. Relative Importance of Defects Detected

In the test situation the examinee has limited time to complete inspection of the vehicles, which had many defects and which, he was told, were about to be put to important use. The likelihood of the vehicles' performing adequately would depend upon the examinee's identification of the more critical faults, such as some of those impairing the functioning of the engine, brakes, and steering mechanism (as compared, for example, with a missing manifold nut or a defective map-compartment catch). It was assumed that a tendency to act in a practical manner to meet the needs of the situation would tend to raise the average level of the importance of the defects detected. Therefore, an importance score was developed as a potential measure of a practical and synoptic approach. Through use of the average importance ratings for the symptom score of each defect, an importance score was developed, separately for each vehicle and for the three vehicles combined, formulated as follows:

> [(importance sum for the N defects detected)-(importance sum for the least important N defects)] divided by [(importance sum for the most important N defects)-(importance sum for the least important N defects)].

This score was expected to have a large chance component, since examinees intent on discovering the most important deficiencies would still observe and record a number of other less significant defects. However, the score was believed likely to contain appreciable systematic variance of the kind intended.

3. Location/Symptom Ratio

A symptom is frequently more immediately apparent than is the fault that is its cause. To identify the latter often requires understanding of underlying physical relationships. The Location/Symptom Ratio was intended to serve as a measure of the understanding of such relationships in automotive functioning, relatively independent of amount of work done (number of defects noted). It represents, among all defects for which the examinee received either symptom or location credit and which were scorable for both, the percentage of credits which were for location rather than for symptom only. The ratio was obtained for each vehicle separately and for all vehicles combined.

4. Making vs Ordering Repairs (Correction Percent)

This score is based on those defects that could be corrected with the equipment available and in the time allowed and that the examinee had fully identified, as evidenced by a location credit. The score is the percentage of such defects repaired. The score was considered likely to have diverse determination, but should often indicate interest in mechanical work, possibly as opposed to supervision or desk work. The score does not take into account the effect of ability level nor the increasing appropriateness of performing more of the mechanical work as the proportion of defects identified approaches unity.

Statistics of Major Variables. Means, standard deviations, and intercorrelations of major variables, including composite scores, were obtained for use in evaluating internal-consistency reliabilities and formulating a total score. In the case of scores obtained separately for each vehicle, reliability of totals across vehicles was estimated from vehicle intercorrelations and also from vehicle and total variances. In addition, reliability of the sum of diagnosis and repair scores was estimated separately for each vehicle from item scores.

Total Score. A final total score was formulated, with a view to comprehensiveness and reliability. In its composition preference was given to the objective scores over the subjective, and to the examinee's accomplishment rather than to scores representing his manner of proceeding.

RESULTS

ITEM STATISTICS

Results of item analyses of the objectively-scored data are discussed in detail in Appendix A. No need was found to eliminate any identifying-information items or vehicle-deficiency items for limited variability, nor any vehicle-deficiency items for poor correlational behavior. Among the vehicle defect items, one intercorrelating cluster consisted largely of disconnected-part defects, and another largely of wiring defects. The four defects proving to be most independent were atypical in content; they related primarily to the vehicle body.

VEHICLE STATISTICS

Table 1 presents, for each vehicle, means and standard deviations of the major scores obtained separately for each vehicle and the between-vehicle correlation of these scores. Means for vehicle 3 tend to be slightly lower for scores representing amount accomplished. (An exception is Made Repairs. The lower mean for vehicle 1 on this variable is at least partially attributable to the relatively small number of defects (7) that could be corrected by the examinee as compared with 11 and 12 on the other two vehicles.) The tendency to larger standard deviations on this vehicle further indicates that some but not all examinees were pressed for time.

RELIABILITY OF ACROSS-VEHICLE SCORES

Table 2 presents for major across-vehicle variables reliability coefficients derived from the separate vehicle scores. The coefficients were obtained by two alternate procedures: first, by application of the Spearman-Brown formula to mean vehicle intercorrelations; second, by Cronbach's generalized internal consistency formula, alpha, through use of vehicle and total variances. Calculated on the same sample, the two measures should be identical when components have equal variances.

No estimates of reliability are available for the various rating variables, which were obtained from only one rater. Concentration vs Distribution of Effort was omitted because data required for computation of its reliability, such as scores based on split halves, were not obtained.

The Identifying Information score has the very high reliability coefficient of .94, attributable in part to identity across vehicles in kinds of information required. Except for Importance of Defects Detected, the remaining variables had generally satisfactory coefficients ranging from the upper .50's to the upper 70's. Variable 8 in Table 2 (most nearly comparable to the total score finally adopted) shows a reliability coefficient of .77. Importance of Defects Detected had reliability only in the low 30's. This variable was not included in the total score. Reliability results are discussed in greater detail in Appendix B.

ANALYSIS OF ACROSS-VEHICLE SCORES

Table 3 presents means, standard deviations, and intercorrelations of major scores, summed across vehicles or otherwise having reference to the entire task. Included are seven Factors Considered. The other three, Keeping Cool, Endurance and Stamina, and Other (to be specified by the rater), were omitted because of small variance and, in the first two instances, because they were not put to test in the task. Included in place of the trial total scores of Tables 1 and 2 is the final task score whose formulation is described below.

Table 1

MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS
OF WORKSHEET VARIABLES AND COMPOSITES FOR VEHICLES 1, 2, AND 3

Variable	Mean			S.D.			r		
	1	2	3	1	2	3	1-2	1-3	2-3
1. Identifying Information	4.4	4.4	4.2	2.2	2.2	2.3	.93	.88	.90
2. Diagnostic Sum	8.0	8.6	7.5	3.8	3.5	4.1	.50	.34	.45
3. Ordered Repairs	1.7	1.5	1.5	1.6	1.5	1.5	.48	.43	.54
4. Made Repairs	.9	1.7	1.2	1.1	1.6	1.4	.49	.40	.55
5. Repair Sum (3 + 4)	2.6	3.2	2.8	2.0	2.0	2.0	.59	.46	.52
6. Diagnosis + Repair Sum (2 + 5)	10.6	11.8	10.3	5.5	5.3	5.8	.54	.37	.47
7. With double wt. for "Made" (4 + 6)	11.5	13.4	11.5	6.2	6.5	6.7	.54	.39	.48
8. Same + Identifying Info. (1 + 7)	15.9	17.9	15.7	6.9	7.0	7.4	.60	.46	.54
9. Importance of Defects Detected	53.1	56.2	43.4	16.2	14.7	18.5	.07	.13	.22
10. Location/Symptom Ratio	.43	.44	.74	.28	.26	.31	.46	.22	.25
11. Correction Percent	54.5	61.9	40.3	41.7	39.3	36.9	.32	.30	.39

Table 2

RELIABILITY COEFFICIENTS

Variable		Spearman-Brown[a]	Alpha[a]
1.	Identifying Information	.96	.96
2.	Diagnostic Sum	.69	.69
3.	Ordered Repairs	.74	.74
4.	Made Repairs	.74	.73
5.	Repair Sum	.77	.77
6.	Diagnosis + Repair	.72	.71
7.	Diagnosis + Repair + Made Repair	.73	.73
8.	Diagnosis + Repair + Made Repair + Identifying Info.	.77	.77
9.	Importance of Defects Detected	.33	.34
10.	Location/Symptom Ratio	.57	.56
11.	Correction Percent	.60	.69

[a] The Spearman-Brown estimate is based on vehicle intercorrelations, the alpha estimate on vehicle variances. The coefficients are subject to certain biases, as described in Appendix B.

Table 3

MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS OF WORKSHEET VARIABLES SUMMED ACROSS VEHICLES AND OF SUBJECTIVE VARIABLES

Variable	Mean	S.D.	I.I.	Diag.	O.R.	M.R.	R.S.	D. + R.	Total	C.E.	Imp.	L.R.	C.%	T.A.	E.A.	Man.	Mot.	Att.	Bear.	Expr.	Fam.	F.I.	Acc.	Cmd.
Identifying Information	13.0	6.8																						
Diagnostic Sum	24.1	9.0	.12																					
Ordered Repairs	4.8	3.7	.03	.50																				
Made Repairs	3.8	3.3	.12	.60	-.00																			
Repair Sum	8.6	5.0	.11	.77	.75	.67																		
Diagnosis plus Repair	32.7	13.3	.12	.97	.62	.66	.90																	
Total Score	500.	100.	.39	.92	.53	.70	.86	.95																
Concentration of Effort	16.5	10.4	-.36	-.27	-.16	-.12	-.21	-.26	-.33															
Importance of Defects	51.8	10.8	.05	.15	-.06	.34	.18	.17	.19	.03														
Location/Symptom Ratio	.51	.21	.07	.59	.24	.70	.65	.65	.65	-.08	.21													
Correction Percent	49.8	31.9	.08	.11	-.26	.74	.30	.19	.26	.05	.29	.42												
Troubleshooting Approach	1.6	1.5	.16	.46	.24	.30	.38	.46	.47	-.14	.06	.28	.07											
Use of Enlisted Assistant	1.7	1.5	.12	.35	.09	.33	.29	.35	.36	-.09	.15	.24	.19	.44										
Use of Manuals	1.8	1.9	.15	.36	.20	.34	.37	.39	.41	-.12	.13	.32	.18	.45	.32									
Motivation	3.1	.96	.12	.53	.30	.35	.46	.53	.52	-.10	.09	.36	.10	.59	.46	.43								
Attitude	3.2	.84	.11	.47	.25	.34	.41	.48	.47	-.11	.13	.31	.12	.51	.45	.35	.80							
Factors: 1. Bearing	1.2	.62	.16	.39	.18	.31	.34	.39	.41	-.14	.11	.28	.12	.49	.40	.39	.57	.58						
2. Expression	1.0	.47	.08	.21	.12	.14	.19	.21	.22	-.04	.06	.16	.03	.27	.18	.25	.31	.34	.31					
5. Familiarity	.59	.75	.09	.47	.21	.38	.41	.47	.47	-.13	.14	.33	.13	.44	.29	.49	.38	.31	.28	.21				
6. Following Instructions	1.0	.72	.15	.26	.13	.21	.24	.27	.29	-.17	.05	.19	.11	.22	.23	.12	.35	.34	.25	.12	.05			
7. Mission Accomplishment	.75	.82	.06	.64	.32	.47	.55	.65	.61	-.21	.11	.41	.15	.49	.36	.43	.49	.43	.37	.18	.46	.20		
8. Effective Command	.92	.48	.12	.17	.07	.15	.15	.17	.19	-.06	.08	.09	.09	.22	.53	.20	.26	.24	.26	.06	.11	.14	.11	
9. General impression	1.0	.38	.08	.24	.07	.22	.20	.24	.25	-.10	.08	.22	.09	.26	.24	.23	.33	.36	.32	.17	.22	.13	.21	.13

Intercorrelations of the four special indices with other test variables provide some evidence whether these indices measure intended examinee characteristics and what else they may measure. The following comments pertain to these intercorrelations.

Concentration vs Distribution of Effort. The largest coefficient is -.5[illegible], with Identifying Information. This negative correlation merely means that those who concentrated on one or two vehicles tended not to make initial entries on the forms for the other vehicles. Representation of thoroughness in this index, as was intended, is indicated by diminishing negative relationships from Diagnosis to Ordered Repairs to Made Repairs. If a defect was detected, individuals high on this index tended more than the average examinee to carry through to appropriate corrective action, especially to the extent of making the repair rather than ordering it. However, the fact that these coefficients are negative (as are those with vehicle totals) indicates that the index also represents lack of ability. For purer measure of thoroughness, correction for ability would be required.

Relative Importance of Defects Detected. This index has positive correlation with all variables representing performance, ability, and other favorable characteristics. (The single negative correlation, with Ordering Repairs, occurs only because examinees high on the index tended to make rather than order repairs.) It appears that here, too, there is an ability component which might have to be removed or reduced if the index is to represent more adequately a tendency toward a practical approach. An interest or attitudinal component, perhaps related to practicality, is indicated by the coefficient of .29 with Correction Percent, the tendency to make rather than order repairs. This .29 is high in view of the estimated reliability of the index (Table 2) of only .33 to .34, and considerably higher than the correlation coefficient of .15 with Diagnosis and .14 with Familiarity with the Task, both of which might be taken as measures of ability in the task.

Location/Symptom Ratio. The highest correlation was with Made Repairs, .70. That location of a defect was a precondition for its correction accounts for the strong relationship. To what extent the systematic part of the remaining variance represents the intended variable of understanding of physical relationships relatively independent of amount of work done and represents it more purely than does Made Repairs is not clear from the statistical data. Among intercorrelational differences between Location/Symptom Ratio and Made Repairs, apparent when allowance is made for the indicated superior reliability of the latter (Table 2), are the weaker relationship of Location/Symptom Ratio with Importance of Defects Detected and the stronger relationship with Motivation. These two differences suggest that the variable may represent in part persistence in tracking down symptoms as distinct from ability to see their significance.

Making vs Ordering Repairs (Correction Percent). The highest correlation was with Made Repairs (.74). However, there were marked differences between the two variables in correlation with other variables. The correlation of Making vs Ordering Repairs was much lower, despite substantial inter-vehicle reliability, with Diagnostic Sum (.11 vs .60), Motivation (.10 vs .39), and Familiarity with Equipment (.13 vs .38). It appears, then, that this variable represents, as intended, preference for making repairs and does not have high ability, effort, or general accomplishment components.

Table 4

IMPORTANCE WEIGHTS GIVEN BY OEC EXAMINERS

Component	Mean Item Weight	Total Weight Raw	Total Weight Percentage[a]
Identifying Information	2.7	60	13
Diagnosis[b]	5.4	227	51
Symptom	3.2	101	(23)
Location	5.4	227	(51)
Corrective Action[b, c]	3.8	160	36
Ordered Repairs	2.6	111	(25)
Made Repairs	4.3	129	(28)

[a] Percentages in parentheses are not additive to the total.

[b] Not all defects have Symptom scores, nor do all have Made-Repair scores.

[c] In cases of alternative corrective actions, the higher weighted alternative was used in computation.

TOTAL SCORE

The total score was formulated to include all components of the objective scoring record except the Identifying Information score for vehicle 3. In arriving at the relative weighting for these components, the importance weights furnished by the OEC examiners were taken into account. Each examiner had given a weight on a 0 to 10 importance scale to each scoring item on the Automotive Inspection Scoring Forms. Table 4 gives, for each main class of scores, the average across items of the average weight given an item, the total of average item weights across items of a class, and the percentage of the grand total of averaged item weights falling in each of the three main classes of scores. In the absence of contrary indications, a weighting of component scores having consistency with these percentages was considered desirable. Additional considerations affecting the selection and weighting of test content were:

Identifying Information. Since the same items of information were to be recorded for each vehicle, consistency across vehicles was high. It was not necessary, therefore, for the score from each vehicle to enter into the total score. Moreover, there was particular reason to omit the score from vehicle 3. An appreciable number of examinees did not start work on vehicle 3 and for these, presence or absence of the identifying information entries would have little significance as representing good or poor performance.

Diagnosis. This appears to be the heart of the inspection task. The respective weights of 1 and 2 given to the Symptom and Location scores are in line with the respective OEC mean importance weights. Reliability as measured was, however, somewhat lower than for Identifying Information and the repair scores.

Repair Scores. The major question concerning repair scores was whether Made Repairs should receive equal or greater score credit in comparison with Ordered Repairs. This matter is discussed in some detail in Appendix C. Consideration of both the assigned mission and the correlational behavior of the two variables led to the decision to weight Made Repairs half again as much as Ordered Repairs.

Subjective Scores. These are the scores for Trouble Shooting Approach, Utilization of Personnel, Use of Available Manuals, Motivation, Attitude, and Factors Considered. What these scores measure, so far as it is important in the Automotive Inspection Task, is likely to affect and be measured by the objective scores. Demonstration of independent contribution to validity would require an external criterion, which may be provided to some extent by certain of the correlation coefficients to be obtained across tests. Meanwhile, there seems little reason to include these scores in the present total score. Moreover, in the case of Factors Considered, some appear irrelevant to the task (e.g., Bearing and Assurance and Expression) and not specific to the technical area, even if otherwise significant; and some in addition appear not to be tested or not readily observed (e.g., Keeping Cool and Endurance and Stamina).

The above considerations led to formulation of a total score consisting of the following components:

1. Identifying Information: raw score on vehicles 1 and 2.

2. Diagnosis: raw score.

3. Ordered Repairs: raw score.

4. Made Repairs: raw score multiplied by 1.5.

The resulting score produced effective weights for these components of 11%, 54%, 13%, and 22%, as shown in Table 5. The composite raw scores were converted to standard scores with a mean of 500 and standard deviation of 100.

The distribution of these scores was somewhat skewed positively. The skew may represent not an artifact but the presence in the sample of a minority of individuals experienced in automotive inspection through Army assignment or having cars as a hobby.

Table [illegible]

COMPONENTS OF THE TOTAL SCORE EXPRESSED AS PERCENT OF MAXIMUM POSSIBLE TOTAL SCORE, PERCENT OF MEAN TOTAL SCORE, PERCENT OF TOTAL-SCORE VARIANCE, AND PERCENT OF TOTAL-SCORE WEIGHT

Component	Percent of Total Score			
	Maximum Score	Mean Score	Variance	Effective Wt. (Var. X Covs.)
Identifying Information	9%	20%	1[illegible]%	11%
Diagnosis	54	56	43	54
Repairs	37	24	38	35
	100	100	100	100
Ordered Repairs[a]	(27)	11	14	13
Made Repairs[a]	(29)	13	25	22

[a] Percents within parentheses are not additive to the Repair percent.

APPENDIX A

ITEM STATISTICS

The p-values of Identifying Information items varied from .51 to .90 for vehicle 1, .[illegible] to .[illegible] for vehicle 2, and .4[illegible] to .[illegible] for vehicle 3.

All vehicle deficiencies that were scored were functional as items. The frequency with which deficiencies or their symptoms were detected and recorded varied from .02[illegible] (a broken lubrication fitting causing leakage) to .940 (an open circuit in a headlight housing causing the light to fail), with a mean of .35. Frequency of location of a defect ranged from .026 to .5[illegible], with a mean of .[illegible]. Frequency of credit for acceptable corrective actions ordered or made varied from .023 to .[illegible], with a mean of .20. Thirty of the 42 defects could be remedied by the examinee. On these 30 items, the proportion of corrective action credits for making repairs rather than ordering them ranged from .05 to .95. In 19 of the 30 items, the proportion was greater than .50. (The proportion .95 was associated with a disconnected distributor primary wire. Connection of this wire is necessary in order to operate the engine as is required for an adequate inspection.)

For the defect scores (on a 0 to 3 scale), standard deviations ranged from .46 to 1.43, the latter being close to the maximum of 1.50 for scores on a 0-to-3 scale. The median was approximately 1.1. The generally high dispersions arose from the tendency for a defect to be completely overlooked or else fully identified (located), with appropriate correction made or ordered. Defect-score intercorrelations ranged from -.10 to +.30. The dozen defect items having one or more high correlation coefficients with other items (.25 to .30) entered into either of two chain-networks, one with strong representation of disconnected-part defects, the other of electrical-wiring defects.

Sixteen percent of the defect item intercorrelations were negative. The number of negative coefficients among the 41 for each item varied from 0 to 22. However, none of the items were judged unsatisfactory on the basis of negative or low intercorrelations. "Competition" among defects within a limited inspection period would be expected to lower the intercorrelations, making slightly negative those coefficients otherwise slightly positive. Also, sampling error would cause some otherwise slightly positive coefficients to become negative. All items had positive coefficients larger in magnitude than their lowest negative coefficients. The items with a large proportion of negative correlation coefficients tend to be less typical in content. Whereas all other items pertain to the engine and related parts, the chassis, and electric-powered accessories, the four items with the largest number of negative coefficients with other items pertain mostly to the body of the vehicle (missing gas-can bracket, inoperative seat adjustment, defective map-compartment door catch, missing publications). These four items form a cluster with intercorrelations of .12 to .23.

APPENDIX B

RELIABILITY OF ACROSS-VEHICLE SCORES

The reliability estimates of Table 2 of the text were affected by certain biases. In the case of Identifying Information, there were identical information requirements for each vehicle, making it likely that an examinee would get credit for all or no items within each set of three corresponding items. The reported coefficients therefore represent to some extent short-time stability, consistency in meeting a particular set of requirements, rather than equivalence, consistency among groups of distinct items representing a common domain. The reported reliability coefficients are overestimates of the latter kind of stability. Consequently, separate vehicle Kuder-Richardson Formula 20 coefficients were calculated for each vehicle. These were, respectively, .81, .91, and .82, still quite high for a small group of items.

The same problem did not exist for measures representing responses to vehicle defects. Defects appear about as diverse across vehicles as within. However, the examinee's freedom to allocate his time among the three vehicles was likely to affect the reported reliability of diagnosis and repair totals across vehicles. Variability among examinees in allocation of time would lower correlation among vehicles, and thus the Spearman-Brown coefficient, and would increase vehicle variances, thereby reducing the alpha coefficient, whereas intrinsic reliability of vehicle scores and reliability of the total (as might be measured against a parallel form of the entire test) were not necessarily decreased.

For the variable Diagnosis plus Repair, the alternate procedure of determining individual vehicle reliabilities (alpha coefficients) and applying the Spearman-Brown formula was followed. The result was a coefficient of .76 (compared to .72 and .71 in Table 2), based on obtained vehicle alphas of .55, .50 and .49. This coefficient may be a slight overestimate, through the effect on vehicle alpha coefficients of the increased vehicle variance caused by the lack of uniformity in time spent on a vehicle. The best estimate on the basis of the available data would then lie intermediate, e.g., at .74. A similar slight increase over the tabled values might thus be obtained for the other total scores representing responses to vehicle defects.

In the case of Making vs Ordering Repairs (Correction Percent), there was a fairly large number of indeterminate vehicle scores (an average of 15% per vehicle) due to a zero denominator (representing absence of any credit for ordering or making repairs) and an even larger number of cases (an average of 26%) missing from vehicle intercorrelations owing to an indeterminate score on one or both the vehicles being corrected. It is likely that many of the examinees with the indeterminate Correction Percent scores were inept or poorly motivated and would therefore have obtained a low Correction Percent score if additional time had been given. If so, the vehicle intercorrelations were reduced through restriction in range, and the Spearman-Brown coefficient derived from them was lowered. Further, the vehicle variances were lowered, and the alpha coefficient derived from these and the total variance (less than 2% of the cases were missing through indeterminate scores) was raised. Had scores been available for the cases with indeterminate scores, the coefficients, as indicated by the above argument, would lie between these two values and nearer the .69.

APPENDIX C

REPAIR SCORES

A consideration in the relative weighting of ordering repairs vs making repairs is the relative merit of the actions themselves in the case of defects the majority) that permitted either action. An argument directed toward lower weighting for making repairs might be that a number of examinees, instructed to undertake repairs but only after discovering as many defects as possible, used time needed for further inspection in making repairs. However, in such cases, the individuals were probably penalized through loss of credit on other defects more than enough to outweigh a sizable differential in credit for making over ordering repairs. Another argument is that making repairs is not a typical officer job. However, this consideration seems irrelevant if making repairs is part of the assigned mission and requires abilities associated with officer performance on other technical jobs. (Most immediately, ability to make repairs might be expected to improve the ability to supervise maintenance, a typical officer job.) A consideration in favor of making repairs is that the activity may assist further inspection, for example, by permitting the engine to run so that other faults in its function may be detected. Also, making repairs represents a larger sample of time and activity, better satisfaction of mission requirements (especially under the circumstances of the OEC scenario), technical interest (since the activity is to some extent optional), and technical abilities supplementing those measured by detection of symptom and location.

A further consideration is the relative magnitude of the ordering repairs and making repairs correlation with the basic task of symptom and location determination, under the circumstance that the reliabilities of the two repair scores are similar. The evidence here is equivocal with respect to determination of satisfactory relative weights. Making repairs had the higher correlation with Diagnosis (.60 vs .50), despite apparently greater difference in required technical abilities. The substantial correlation of making repairs, while evidence of its relevance, indicated that more of its systematic variance is accounted for in the Diagnostic Sum score than that of ordering repairs. On the other hand, it is possible that the unaccounted-for variance of ordering repairs has some relatively unimportant components representing, for example, clerical follow-through in completing the worksheets. If so, provision of greater weight for ordering repairs would not be appropriate on account of the greater independent variance.

However, in addition to the fact that making repairs correlated more highly with Diagnostic Sum than did ordering repairs, it also correlated more highly with the whole set of judgmental ratings beginning with Trouble-shooting Approach (Table 3). The decision was therefore made to weight Made Repairs more than Ordered Repairs, but by half again as much rather than twice as much (the latter being the weight used in variables 7 and 8 of Tables 1 and 2). Thus, the weight of 1.5 for Made Repairs was used in computing Total Score as finally constituted.